第29卷第2期 中 国 科 学 技 术 大 学 学 报 Vol. 29, No. 2
1999年4月 JOURNAL OF CHINA UNIVERSITY OF SCIENCE AND TECHNOLOGY Apr. 1999

A Method of Data Expansion Based on Wavelet Transform*

SHAO Limin SHAO Xueguang

(*Department of Chemistry, USTC*)

Abstract A method of data expansion is established in this paper based on wavelet transform. Numerical experiments are carried out to test the performance of the method. In comparison with the result of cubic spline interpolation, it shown that the expanding error of this method is much smaller than that of cubic spline interpolation. It is also found that, with the increase of the original data length, the expanding error of this method decreases more sharply than that of cubic spline interpolation. As an application, a signal of High Performance Liquid Chromatography (HPLC) is processed by the method, of which the result is satisfactory.

Key words Data expansion, Wavelet transform, Cubic spline interpolation

1 Introduction

Wavelet transform is a new mathematical technique, which is widely applied in several fields such as signal analysis, image processing and pattern matching, etc. Applications of wavelet transform in analytical chemistry are being explored in recent years, results of which show advantages of the technique[1,2,3,4]. Discrete wavelet transform (DWT) is generally used in practice, and the most important algorithm in DWT is Mallat Pyramid. By Mallat Pyramid algorithm a discrete signal will be decomposed into a series of contributions, and each contribution has half data length of the previous one. In the reconstruction of the original signal from these contributions, the data length after each calculation will be doubled until the original signal is restored.

In this paper, a method of data expansion is established based on the reconstruction algorithm of the Mallat Pyramid. The reconstruction was employed to treat a signal rather than the decomposed contributions. Its result means data expansion.

Theory

* 收稿日期: 1998-02-27
邵利民: 男, 1972年7月生, 博士研究生. 邮编: 230026, 合肥

.1 Theory of Discrete Wavelet Transform

Let $A_{2^j}^{d_j} f$ and $D_{2^j}^{d_j} f$ denote the 'discrete approximation' and 'discrete detail' of function (or signal) $f(x) \in (V_{2^j})_{j \in Z}$ at the resolution 2^j ($(V_{2^j})_{j \in Z}$ is a multiresolution approximation of the Hilbert space). h(n) and g(n) are used to denote the low pass filter and high pass filter. Mallat proved[5] that there existed recurrence formulae.

$$A_{2^j}^{d_j} f(n) = \sum_{k=0}^{N-1} h(k-2n) A_{2^{j+1}}^{d_{j+1}} f(k) \tag{1}$$

$$D_{2^j}^{d_j} f(n) = \sum_{k=0}^{N-1} g(k-2n) A_{2^{j+1}}^{d_{j+1}} f(k) \tag{2}$$

where $-J \le j \le -1 (j \in Z, j > 0)$. J is a preset scale parameter, and it is also the maximum decomposing time. N is the data length. When $j = -1$, the discrete approximation $A_{2^{j+1}}^{d_{j+1}} f = A_{2^0}^{d_0} f$ is the original signal.

Decomposition of a signal by formulae (1) and (2) is the Mallat Pyramid algorithm. By this algorithm the original signal will be decomposed into discrete approximation $A_{2^j}^{d_j} f$ and discrete detail $D_{2^j}^{d_j} f (-J \le j \le -1)$ with j decreasing from -1 to $-J$. What is important is that the data length of $A_{2^j}^{d_j} f$ (or $D_{2^j}^{d_j} f$) is $2^j \cdot N$ while the data length of the original signal is N. Thus a discrete signal can be compressed by the Mallat Pyramid. Furthermore, because $\boldsymbol{H}(=\{h(n)\})$ is a low-pass filter and $\boldsymbol{G}(=\{g(n)\})$ is a high-pass one, $A_{2^j}^{d_j} f$ and $D_{2^j}^{d_j} f$ respectively represent the low-frequency and high-frequency information contained in the original signal.

The original signal can be reconstructed from $A_{2^j}^{d_j} f$ and $D_{2^j}^{d_j} f$, and the data length of the original signal can also be restored by the following process:

$$A_{2^{j+1}}^{d_{j+1}} f(n) = \sum_{k=0}^{N-1} h(n-2k) A_{2^j}^{d_j} f(k) + \sum_{k=0}^{N-1} g(n-2k) D_{2^j}^{d_j} f(k) \tag{3}$$

This process will be repeated with j increasing from $-J$ to -1 until $A_{2^0}^{d_0} f$, the original signal, is reached.

. Theory of the Data Expansion Based on Wavelet Transform

From the introduction above, we can see that $A_{2^{j+1}}^{d_{j+1}} f$ has double data length after being reconstructed from $A_{2^j}^{d_j} f$ and $A_{2^j}^{d_j} f$. Therefore, if the reconstruction algorithm is used to process a signal, the result is that the data length of the signal will be doubled. The data length can be increased to 4, 8, ..., 2^N times of the original one merely when the reconstruction is repeated 2, 3, ..., N times. Furthermore, in decomposition by wavelet transform, it had been proved[5] that in comparison with the other kinds of discrete approximations at resolution 2^j, $A_{2^j}^{d_j} f$ is the most similar to the original signal $f(x)$. So, let $f(n)$ be a discrete sample of function $f(x)$ with data length being N. Then at resolution 2^j, equation (4) can be derived

$$A_{2^j}^{d_j} f = f(n) + o(N) \qquad (1 \le n \le N) \tag{4}$$

At resolution 2^{j+1}, the data length of the discrete sample $f(n)$ will be doubled. From (4), there will be:

$$A_{2^{j+1}}^{d}f = f(n) + o(2 \times N) \qquad (1 \leq n \leq 2 \times N) \tag{5}$$

Formula (5) means that of all the discrete approximations at resolution 2^{j+1} with data length being $2 \times N$, $A_{2^{j+1}}^{d}f$ is the most similar to the original signal, which ensures the accuracy of the data expansion. In the implementation of the data expansion, $D_{2^j}^{d}f$, the discrete detail, is set zero in order to avoid the introduction of high frequency noise. Thus the formula of data expansion based on wavelet transform can be written as

$$A_{2^{j+1}}^{d}f(n) = \sum_{k=0}^{N-1} h(n - 2k) A_{2^j}^{d}f(k) \tag{6}$$

The data expansion algorithm can be illustrated by the following scheme:

$$\begin{array}{ccccccccc} A_{2^J}^{d}f & \rightarrow & A_{2^{J+1}}^{d}f & \rightarrow & A_{2^{J+2}}^{d}f & \rightarrow \cdots \rightarrow & A_{2^1}^{d}f & \rightarrow & A_{2^0}^{d}f \\ & \nearrow & & \nearrow & & \nearrow \cdots & & \nearrow & \\ D_{2^J}^{d}f & & D_{2^{J+1}}^{d}f & & D_{2^{J+2}}^{d}f & \cdots & D_{2^1}^{d}f & & \end{array}$$

Data length: $N \quad 2^1N \quad 2^2N \quad \cdots \quad 2^{J-1}N \quad 2^JN$

Where $A_{2^J}^{d}f$ denotes the original signal to be expanded and $A_{2^0}^{d}f$ denotes the result of the expansion. It is obvious that after J times calculation, data length was expanded to be 2^J times of the original one ($J \in Z$, $J > 0$).

3 Numerical experiments

In this section, the performance of the data expansion method is tested with numerical experiments. Because many signals have some characteristics of the Gaussian curve, the original signal chosen for the test is the Gaussian peak. Function of the signal is given as:

eq $$y = \exp[-(x - 50)^2/100], \quad 30 \leq x \leq 70 \tag{7}$$

The expanding performance is judged by Root Mean Square Difference(RMSD) that is calculated by

$$E = \sqrt{\frac{\Sigma(X_i^P - X_i^T)^2}{N}} \tag{8}$$

Where X_i^P is the calculated value by the data expansion, X_i^T is the theoretical value, and N is the data length of the expanded signal.

For comparison, the same data expansions were also performed with cubic spline interpolation.

3.1 Determination of wavelet

Several wavelets have already been found since the establishment of wavelet transform theory, and each has some peculiar properties. In order to determine the most suitable wavelet for data expansion of Gaussian curve, discrete signal of function (7) is prepared. The original data length is set 100, and will be expanded to 200.

Five wavelets were used to perform data expansion. They are Daubechies (Filter length: 4, 6, 8, 10, 12, 14, 16, 18, 20), Beylkin (Filter length: 18), Coiflet (Filter length: 6, 12, 18, 24, 30),

Symmlet (Filter length: 8, 10, 12, 14, 16, 18, 20) and Vaidyanathan (Filter length: 24). According to values of RMSD, it is found that the Coiflet wavelet with filter length 18 can obtain the most accurate expanded signal in the case of the original signal being Gaussian.

3. Test 1

The data length of the original signal obtained from function (7) is preset 100. After the data expansion, the data length is increased to 200, 400, 800 and 1600. The errors of data expansion with either wavelet transform or cubic spline are tabulated in Table 1.

Tab. 1 Comparison of performance between wavelet transform and cubic spline interpolation with different expanding multiples

Length of the expanded signal	Expanding Error (RMSD)	
	Wavelet transform	Cubic Spline
200	3.53141×10^{-6}	1.78485×10^{-3}
400	3.83199×10^{-6}	1.78841×10^{-3}
800	4.04293×10^{-6}	1.66889×10^{-3}
1600	4.16600×10^{-6}	1.78849×10^{-3}

From Table 1, it can be found that the expanding errors generated by wavelet transform are much smaller than those by cubic spline at every expanding multiple. Even when the data length is increased to 1600, which means the expanded signal has 8 times data length of the original one, the expanding error of wavelet transform is as small as 4.166×10^{-6}, whereas the expanding error of cubic spline is 1.788×10^{-3}.

The results show that data expansion with wavelet transform can perform high multiple expansions with little loss of accuracy.

3. Test

The performance of the data expansion method based on wavelet transform is tested in the original signals have ditterent data lengths.

Tab. The comparison of the expanding errors generated by wavelet transform and cubic spline with different original data length

Original data length	Expanding Error (RMSD)	
	Wavelet transform	Cubic Spline
50	2.03452×10^{-5}	2.45990×10^{-3}
100	3.53141×10^{-6}	1.78485×10^{-3}
150	1.27382×10^{-6}	1.46992×10^{-3}
200	6.18676×10^{-7}	1.27828×10^{-3}
250	3.53517×10^{-7}	1.14647×10^{-3}
300	2.23841×10^{-7}	1.04838×10^{-3}
350	1.52126×10^{-7}	9.71809×10^{-4}

The original signals are also obtained from function (7), and the data length varies from 50 to 350 with step 50. For simplicity, the signal is expanded to double data length of the original one. The expanding errors that are also calculated by formula (8) are tabulated in Table 2.

From Table 2, it can be found that at every data length, the expanding error of wavelet transform is much smaller than that of cubic spline interpolation, which is consistent with the results of test 1. In order to make the comparison more clear, Figure 1 is drawn according to values of Table 2. From the figure we can see that with the increase of the data length of the original signal, the expanding error of wavelet transform (or cubic spline interpolation) decreases. This is because increase of the original data length means the discrete signal con-

tains more information of the prototype function. However, the expanding error of wavelet transform decreases more sharply than that of cubic spline interpolation. This means that the longer the data length of the original signal is, the more accurate the expanded signal will be obtained by the data expansion with wavelet transform rather than with cubic spline interpolation. In other words, if a discrete signal is measured with a few more data points than before at some cost, with wavelet transform a more accurate expanded signal will be obtained.

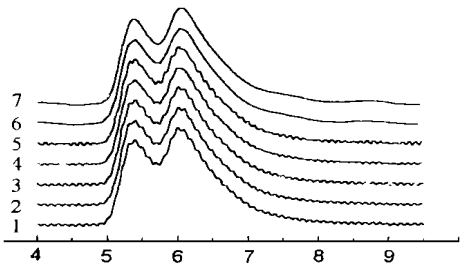

Fig. 1 Logarithmic values of expanding errors via different data length of original signal
(‘ o’ wavelet transform, ‘ •’ cubic spline interpolation)

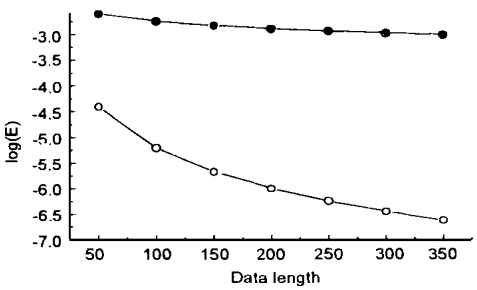

Fig. Comparison of expanded signal and the experimental chromatogram

4 An application

As an application, a signal of High Performance Liquid Chromatography (HPLC) without a perfect Gaussian shape is processed by the data expansion based on wavelet transform.

A sample for HPLC is the solution of Yb and Tm of which the concentrations are 20.01$\mu g \cdot mL^{-1}$ and 19.99$\mu g \cdot mL^{-1}$. The chromatogram of the sample is measured by a Spectrasystem FL 2000 (Spectra-Physics Inc., U.S.A) with sampling interval being 0.012 min, and its data length is 480. In order to test the performance of the method, three signals are obtained by keeping one out of 2, 4 and 8 points of the chromatogram, which simulate signals measured with sampling interval being 0.024, 0.048 and 0.096 min. The data lengths of the signals are 240, 120 and 60. The data expansion based on wavelet transform is employed to expand the signals to data length of 480. The results are illustrated in Fig 2.

From Figure 2, we can see that both peak height and peak position of the three expanded signals are consistent with those of the experimental chromatogram. This means that there is almost no loss of chromatographic information after the data expansion with wavelet transform. However, there is some difference between the expanded signal (7) and the experimental chromatogram in Figure 2. This might be a result from the faer that the original signal (6) was obtained by keeping one out of 8 points of the chromatogram, which results in some loss of chromatographic information, especially of the noise. However, the main characteristics are maintained. So it is possible that wavelet transform can expand a signal sampled in a short span of time to meet the purpose of scientific research, which can surely save time and energy with least detriment to accuracy.

5 Conclusion

From test 1, test 2 and the application, we can conclude that wavelet transform shows advantages over cubic spline in data expansion. Further more, unlike the cubic spline interpolation the method needs no initial conditions, which is very convenient for calculation. If a signal obtained from an instrument has insufficient data because of some limitations or for the purpose of quick sampling, wavelet transform can be employed to expand the signal.

It should be pointed out that the expansion of the method is 'inter' which means that the expansion does not change the original sampling range.

References

[1] Bos M, Hoogendam E. Wavelet transform for the evaluation of peak intensities in flow injection analysis. Anal. Chim. Acta, 1992, 267: 73—80

[2] 邵利民, 邵学广, 唐兵, 刘善堂. 小波变换用于高效液相色谱的噪声滤除. 分析化学, 1997, 25(1): 15—18

[3] 潘忠孝, 邵学广, 仲红波. 小波变换用于高效液相色谱的基线校正. 分析化学, 1996, 24(2): 149—153

[3] Chau F T, Shih T M, Gao J B, Chan C K. Application of the fast wavelet transform method to compress Ultraviolet-Visible spectra. Applied Spectroscopy, 1996, 50(3): 339—348

[5] Mallat S G. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1989, 11(7): 674—693

一种基于小波变换的数据扩展方法

邵利民 邵学广

(中国科学技术大学化学系)

摘要 提出了一种基于小波变换的数据扩展方法. 数字实验结果表明: 这种方法的误差远小于三次样条插值的结果. 而且随着原始数据点数的增加, 这种方法的计算误差下降得也较三次样条的结果为快. 从而说明了这种基于小波变换的数据扩展方法的准确度是很高的. 作为一个实际应用, 处理了一个的高效液相色谱图, 结果是令人满意的.

关键词 数据扩展, 小波变换, 三次样条插值

中图法分类号 O241.3, O652.7